Data Errors, Structural Change, and Time Series Shocks in the Electricity Market

by

Lindolfo Pedraza, Science Applications International Corporation
Nancy Kirkendall, U.S. Department of Energy
Joe Sedransk, U.S. Department of Energy

Lead Author:

Lindolfo Pedraza
Senior Statistician/Economist
Science Applications International Corporation
901 D Street, SW, Suite 201
Washington, DC 20024

Ph: 202/488-6611
Fax: 202/488-3158

Lindolfo.pedraza@saic.com

OPENING

This paper has two objectives, sharing the results of on-going evaluations of survey sampling and imputation methodologies and collecting feedback and suggestions for refining existing results and implementing future work. We are using publicly available electricity generation micro-level data and current imputation methods to test if data cleaning techniques and microeconomic theoretical concepts can be used to:

1. Improve the accuracy of currently used survey sampling and imputation methods; and
2. Improve the consistency of the time series processes these survey aggregates represent.

In addition to this, we are evaluating how to use and compare alternative imputation models and parameter estimates across sample years to test:

1. Alternative stratification and estimation models; and
2. Alternative estimation methods.

The optimal model specification derived from the second set of bullets is then used to test if structural change in the electric industry can be characterized via the imputation model's estimation parameters. Some preliminary examples are shown in part II below.

While monitoring real-time data, in theory, can prevent short-term market trends that are unsustainable in the long-term [1], monitoring mid- and long-term trends is necessary for the prevention of wasteful over-investment of scarce resources or unsustainable long-term trends of high prices due to chronic under investment.

This paper proposes techniques for how to best analyze and prepare input data used in the estimation of state-level electricity demand estimates (Part I), and alternative ways to improve the estimation accuracy of implemented models (Part II). Part I and an introduction to part II were presented during the 25th annual North American Conference of the USAEE/IAEE. The current version of the paper benefits from additional work in part II and constructive lessons and comments received during the North American Conference of the USAEE/IAEE and from a referee representing the ASA Committee on energy statistics.

Part I describes two sections of our analysis. First, we show how data preparation techniques have a profound effect on the consistency of time series data. Second, we show how microeconomic theory could be inserted into survey sampling and imputation to:

1. Improve the accuracy of monthly estimated of aggregated demand; and
2. Improve the consistency of time series data.

Part II shows preliminary work in the implementation of cross-sectional modeling into survey sampling and imputation. Three closely intertwined statistical and economic issues motivate this effort:

1. Differentiating between regional seasonality and economic growth;
2. Stratification analysis; and
3. Interest in defining if there has been a palpable effect on the electric industry after deregulation in some states.

[1] For example: Analysis of California's transaction data real time concluded on changes to the California trading rules to avoid looping trade.

The main datasets in the study are:

- Published data from the U.S. Department of Energy survey form EIA-826;
  - State level electricity sales (Electricity demand)
  - State level electricity revenues.
- Reported individual data from the EIA-826 and EIA-861 survey forms;
  - State level electricity sales (Electricity demand)
  - State level electricity revenues.

PART I

INTRODUCTION

As the need for improved forecasting efficiency grows, so does the need for consistent and more accurate data to feed such forecasts. Failure to produce clean data for reliable load forecasts could lead to erroneous expectations of future economic developments and, as a result, mistaken and expensive investment decisions.

While consistency is a desired property of time series data, accuracy is a desired property of survey sampling estimates. This paper works with the implied links between economics, statistics and econometrics and brings them explicitly into the analysis arena.

Data cleaning techniques are used to prepare data for publication; data cleaning refers to the systematic use of listings and multiple interactive and drill-down plots to identify erroneous data entries that would otherwise bias survey sampling estimates and damage time-series consistency.

From a statistical stand-point, accurate estimates are those which are not biased by erroneous survey form entries (dirty data). From an economic stand point, un-accurate survey sampling estimates could be erroneously assumed to result from time series shocks due to a structural change in the economic landscape. Our research seeks to place these relationships in perspective.

The data preparation techniques include the linking of cross-sectional and time series modeling. In theory, improvements to intra-temporal or cross-sectional modeling should improve inter-temporal data generation processes of state and regional electricity sales and revenue estimates. Examples that illustrate such potential improvements are provided for the states of Arizona, Delaware and Kansas, to name a few.

Intra-temporal modeling refers to modeling undertaken within a sample period and is better known as cross-sectional modeling. The imputation systems used in U.S. Department of Energy survey forms EIA-826 and EIA-861 are examples of cross-sectional modeling[2]. Inter-temporal refers to "across sample periods" and is better known as time series. The historical data lines of revised EIA-826 and EIA-861 estimates, dating back from their inception until today, are examples of time series.

Co-integration is a time series technique used to estimate short- and long-term relationships among modeled variables. Instead of only using information found in the long-term span of a variable to determine statistical inference, co-integration uses both short- and long-term information to model long-term forecasts as a function of short-term adjustments and long-term trends. Co-integration uses first and second moments information found in the targeted variables. These refer to the mean, median and

[2] Please refer to published EIA-826 for an example. The methodology underlying the current system was developed by James Knaub, EIA, and is documented in his many references

variance of a variable's levels and differences. For example, these differences can be in terms of month-to-month variations.

Modern, more accurate and efficient forecasting techniques such as co-integration require that data not only describe each sample period's cross-sectional level value and some overall long-term trend but also consistently describe the time-series path as a series of short-term relations. In other words, several cross-sectional observations or months can only be consistently linked in a time-series process if they are pooled from the same universe and are based on an accurate cross-sectional imputation model.

This paper uses a 'general-to-specific approach' to data analysis and great data detail, and tries to separate data driven shocks from economic ones. It attempts to shed some light onto the sources of the modeling and survey outcome disagreements, whether they are attributed to data errors or failed expectations.

DATA DESCRIPTION

The subject data are the published and revised EIA-826 (described in detail below) estimates. In a given year, EIA-826 and EIA-861 (also described in detail below) data are cleaned and revised. EIA-826 data are then adjusted to the EIA-861 state level by multiplying each EIA-826 state level estimate by its corresponding state level EIA-861/EIA-826 ratio.

The form EIA-826, "Monthly Electric Sales and Revenue with State Distributions Report," collects information from electric utilities, energy service providers, and distriution companies that sell or deliver electric power to end users. Data collected on this form include sales and revenue for all end-use sectors (residential, commercial, industrial, and transportation). The Form EIA-826 is to be completed by those electric utilities, energy service providers, and distribution companies that sell or distribute electric power to end users and have been selected to report electric energy information on a monthly basis. The set of respondents for the Form EIA-826 is a cut-off sample[3] of respondents chosen from the respondent frame of the Form EIA-861, "Annual Electric Power Industry Report."

Form EIA-861, "Annual Electric Power Industry Report," collects information on the status of electric power industry participants involved in the generation, transmission, and distribution of electric energy in the United States, its territories, and Puerto Rico. The data collected on this form are used to monitor the current status and trends of the electric power industry and to evaluate the future of the industry. The EIA-861 is a census of the power industry companies on the frame.

Data will be reviewed in detail from 2000 and forward; gained insights will then be applied to the whole series starting in 1990. With this, we wish to first isolate statistical noise due to recent deregulation exercises and then compare if the performance of current imputation techniques changes due to de-regulation.

SUMMARY OF IMPUTATION SYSTEMS

A sample of the EIA-861 universe is required to report sales and revenues levels on a monthly basis through form EIA-826. After the data are submitted over EIA's Internet Data Collection (IDC) system, a group of data analysts compares the new numbers with previous ones via interactive analysis tools in order to catch and correct data errors before publication. Once this is completed, the submitted data are sent to the imputation system.

[3] It is currently assumed that cut-off sampling performs better for skewed populations with a few large and many hard to reach small electricity producers such as the population of electricity generators in the United States.

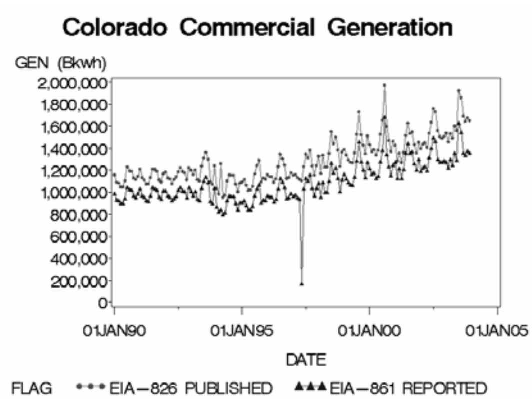
The imputation system[4] for the EIA-826 is used to provide estimated values for non-sampled EIA-861 firms as well as incomplete responses and values that fail editing. See graph 1 below for an example.

The system itself utilizes the most recently finalized EIA-861 data as a regressor value against currently reported EIA-826 monthly data. Reported state level values are subsetted into stratification regional groups for estimation purposes. Weighted regressions are applied to each estimation group, and estimates for each non-sampled EIA-861 observation (at the plant and state level) are produced.

If a firm does not respond to the EIA-826 monthly survey, or if their reported value is deemed questionable, then the estimated values are used in place of the reported ones. If a plant is not required to report EIA-826 data, then their estimated values are also used.

Graph 1: Colorado Commercial Sales or Generation shows the published estimates in red and the reported data in blue. The spread between both lines is the area estimated for by the imputation system. See how the spread is constant with the exception of a single data point, where one or more non-respondent's missing values were resolved by the imputation system.

Graph 1: Colorado Commercial Generation



ANALYSIS TOOLS DESCRIPTION

The main analysis tools are interactive plots that allow analysts to select individual points in a plot to obtain specific information from that point. In addition to this, the analyst can eliminate the case of the plot, the plot will then re-adjust itself to allow the annalist to see other data points that would otherwise be visually unavailable. In addition to this, a drill-down plot system was designed and is currently being implemented to analyze each of the EIA-826 and EIA-861 datasets and identify what monthly estimates need to be reviewed first.

This drill-down plot system allows the user to view and interact with a data analysis outputs to review specific data summaries and plots for a given occurrence of the cross-sectional model. The summary tables and plots are broken up by sector and schedule and contain several interactive features. When the user clicks on one of the points in the plot, they are 'drilled' down further into the data, where summary tables and plots are displayed.

The drill-down plot system will incorporate the use of the analysis tools below:

[4] The methodology incorporated into the imputation system was developed by James Knaub and is documented in his extensive list of references.

- Inter-temporal data manipulation as the source of time series plots.
    - Drill-down plots starting on revised eia-826/861 estimates
    - Monthly and annual dummy parameter data base
    - Listings of EIA-861 greatest differences
        - Levels
            - EIA-861 Year vs. Year
            - EIA-826 Month and Year vs. Month and Year
            - EIA-826 Total Month and Year vs. Total Month and Year
        - Percent Change
            - EIA-861 Year vs. Year
            - EIA-826 Month and Year vs. Month and Year
            - EIA-826 Total Month and Year vs. Total Month and Year
    - Listing of EIA-826 greatest differences
        - Same Levels and Percent Change as Above
    - EIA-826 Scatterplots
        - Levels vs. levels scatterplots are used to compare two reported values from the same respondents.
        - Percent change vs. level change Scatterplots are used to analyze individual percent changes while normalizing for the magnitude of the change itself.
        - Level change vs. change market share scatterplots are used to analyze individual level changes while normalizing for the effect the change has on the total estimate.
    - EIA-861 Scatterplots
        - Levels vs. levels scatterplots are used to compare two reported values from the same respondents.
        - Percent change vs. level change Scatterplots are used to analyze individual percent changes while normalizing for the magnitude of the change itself.
        - Level change vs. change market share scatterplots are used to analyze individual level changes while normalizing for the effect the change has on the total estimate.
- Use of EIA-861 Imputation to resolve visual outliers and missing observations.
- Use of EIA-826 Imputation to resolve visual outliers and missing observations.

ANALYSIS OUTLINE

The first step in the analysis process was to recreate EIA-826 published data in a single imputation method. Up to this date, only observed data and total estimates have been available, and different imputation methodologies have been used in different survey years, making cross-comparison difficult. No de-aggregated imputed data are available, and previous imputation systems[5] are not fully documented. Luckily, recreated EIA-826 time series data were significantly similar to published estimates although slight differences were persistent.

The second step in the analysis of historical data was the visual comparison of the new EIA-826 published data to identify areas where contradictory EIA-861/EIA-826 ratios exist. Future work will

[5] While imputation methodologies have been extensively documented by James Knaub, imputation system documentation has not been fully documented from year to year.

make use of the monthly and annual dummy parameter data base to identify only those plots with significant contradictions in the EIA-861/EIA-826 ratios.

The objective in the second step was two-fold: to provide consistent detailed imputation and prediction data outcomes currently unavailable, and to see if the use of different imputation methods across years account for some of the observed shocks in published estimates. The resulting data shows that published EIA-826 data is consistent with the new modeled data and that the use of different imputation methods across years has not, up to this point, been the source of the observed shocks in published estimates.

The EIA-826 imputation system created new estimates in the following ways:

- For each month and year in the study, we ran the finalized EIA-826 imputation system using finalized previous year EIA-861 data as the regressor data[6].
- Each state monthly EIA-826 estimate will be multiplied by the $EIA861 \Big/ \sum_{1}^{12} EIA826$ ratio to create the new EIA-826 data lines.

The third step was to detect outliers and influential observations in the universe data in EIA-861 and to replace these outlying observed values with predictions based on the current EIA-826 imputation model. The step's objective was to test if some of the time-series inconsistencies in EIA-826 were not time-series shocks of an economic nature.

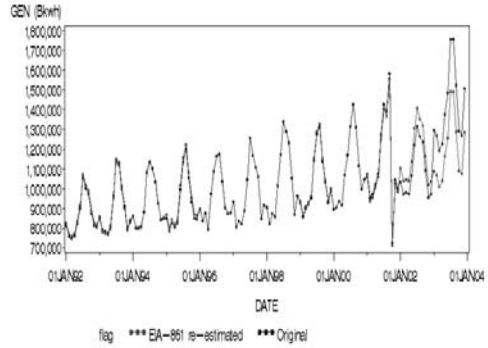
The fourth step was to run the imputation model from step one with the new regressor data outlined in step three. Drill-down plots were used to inspect the state estimates.

The results were positive overall. Of the 51 states analyzed, Arizona's 1999 industrial sector data, Delaware's 2001 and 2002 commercial sector data, Idaho's 2004 commercial and industrial data, Oklahoma's 2004 commercial data, Indiana's 2001 commercial data, Kansas' 1996, 1998 and 2001 industrial sector data, North Dakota's 1995 and 2001 commercial data, Nevada's 1998 and 1999 industrial data, Oregon's 2000 industrial data and Washington's 2000, 2001 and 2002 industrial data showed improvements.

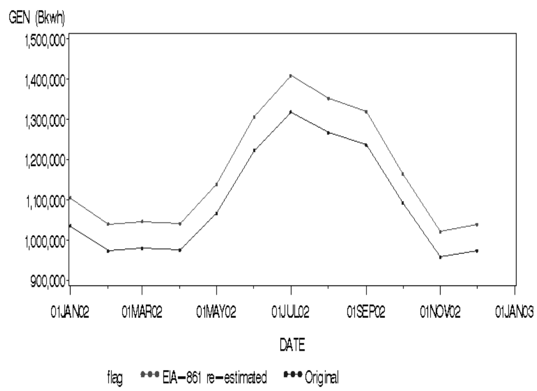
Graphs 2, 3 and 4, below, show Commercial sector electricity demand estimates. The blue line represents published estimates while the red line represents the alternative demand estimates as outlined in step two. Graph 2 shows the long term trend wile graphs 3 and 4 Drill down into the year 2002 and 2003 to closely inspect the new trends.

Graph 2: Oklahoma Electricity Sold to the Commercial Sector

[6] In practice, regressor data is only used after it has been finalized; as a result, two-year prior data is used to impute and publish monthly preliminary estimates. Finalized EIA-826 data is re-imputed Once one-year prior data is finalized.

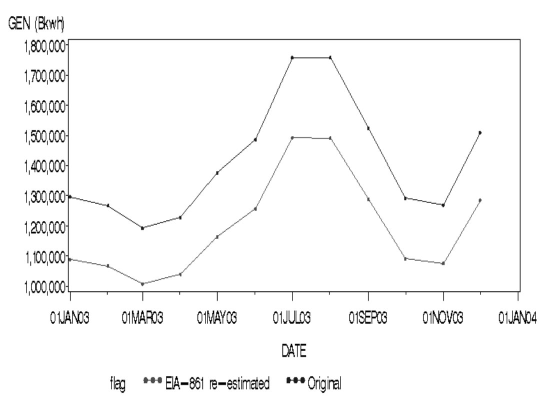


Graph 3: 2002 Oklahoma Electricity Sold to the Commercial Sector



One effect is displayed in graph 3. The new EIA 861/EIA 826 ratio changes the mean of the EIA-826 data. Drills down tools attribute this to change to the re-estimation of two observations. However, only the market share of one of them changes considerably from 38.44% to 41.16%. The second firm's market share changed from 38.74% to 37.46%.

Graph 4: 2003 Oklahoma Electricity Sold to the Commercial Sector



Two effects are displayed in graph 4. On one hand, the effect from the EIA 861/EIA 826 ratio changes the mean in EIA-826; on the other hand, improved 2002 (see graph above) regressor data has also an effect on the overall seasonal pattern. The difference between both lines is of roughly 200,000 in January and 300,000 during the summer. According to the drill down system, two observations changed during 2003.

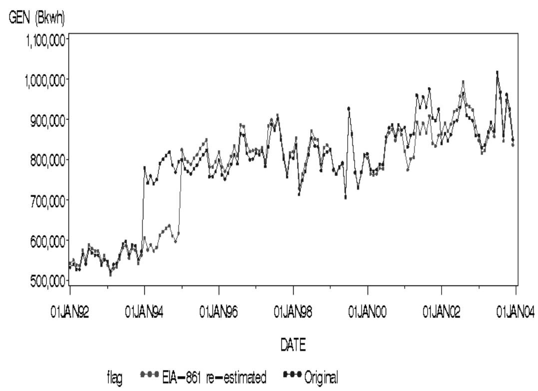
One saw its market share changing from 44.91% to 39.07% while the second one changed from 36.27% to 38.78%.

While the level change seen in graphs 3 and 4 is attributed to changes in the same year EIA 861/EIA 826 ratios, the monthly net change in graph 4 is attributed to changes in the 2002 regressor data. This indicates that EIA-861 data has two important effects on EIA-826 data. EIA-861 not only provides with a frame bench mark but it also provides information with which mid-term trends are estimated.

In addition to this, Colorado's 1994 industrial sector data, Maryland's 1995 commercial and industrial data, Montana's 2001 and 2002 industrial data, New Jersey's 2000 industrial data, Pennsylvania's 1999 and 2000 industrial data, South Dakota's 2001 industrial data and Tennessee's 1999 commercial data show discrepancies or information lags between survey forms EIA-826 and EIA-861. These discrepancies are characterized by lags in the perceived shocks. For example, while currently published data shows a large shock from December 1993 to January 1994 alternative estimates show how this shock is "transferred" happens from December 1994 to January 1995.

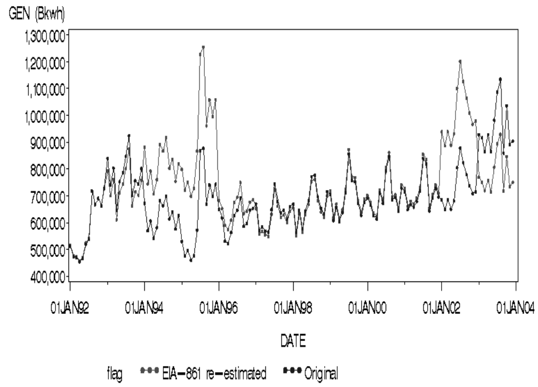
Graph 5 shows sales electricity to the industrial sector in Colorado. While the EIA-861 imputation system "corrected" this trend in 1993, observed 1994 EIA-826 confirmed the correction was not necessary. Although, this least how we currently interpret the results additional analytical work is needed to confirm these differences.

Graph 5: Colorado Electricity Sold to the Industrial Sector



Iowa, New Jersey, Ohio and Texas' data showed less consistency after applying the new imputation models. For example, graph 6 shows sales of electricity to the industrial sector in Iowa. The use of the EIA-861 imputation system creates considerable changes to the time-series data during 1994, 1995, 2002 and 2003. Contrary to the improvements exemplified in graphs 2 through 5, application of these methods, do not provide with smoother seasonality processes through out. However, although 1994, 1995 and 2002 monthly estimates do not keep with "trend" the 2003 trend was improved. As a result, the results in Iowa, New Jersey, Ohio and Texas are mixed.

Graph 6: Iowa Electricity Sold to the Industrial Sector



The fifth step of our analysis focused on the detection and treatment of outlier and influential observations in EIA-826 monthly data. Two alternative models were compared. The first model considers outlier and influential observations as erroneous data and overwrites the values with imputed ones as in the third step. The second model considers outlier and influential observations as independent to the imputation model and extracts them off the imputation model but includes them for estimation proposes.

While the first model simply assumes incorrectly reported data, the second assumes that though the data may be correct, they do not represent the behavior of the smaller companies. Hence, the second model incorporates microeconomic theory to adjust the imputation model. Mainly, the imputation model assumes outlier and influential observations to be oligopolistic competitors. That is, large firms have a competitive advantage against smaller firms. As a result, the smaller firms choose not to compete with the larger more competitive firm and wait until the large firm reaches its capacity. Instead, small firms will move to compete for the residual demand with firms of similar size.

The results are mixed in this case. Both models represent two extreme positions. While the first model tends to flatten seasonal patterns, the second fails to correct for missing observations or grossly erroneously reported values.

Two conclusions are taken from this. First, detection of outlier and influential observations and automatic treatment of those observations, while useful and illustrative, does not provide a global solution to prepare pre-publication data. However, the use of automatic detection tools can greatly improve the performance of other analytical tools such as interactive scatterplots by pointing analyst towards those estimates of greatest importance. We are currently seeking alternative statistical constructs to continue testing automatic outlier detection tools to aid in the implementation of interactive scatterplots.

PART II:

ESTIMATION MODEL ANALYSIS

While the first part of this study evaluated the overall state of existing data and the performance of current imputation models, the second part of this study focuses on the fundamental theory under which current imputation methods are based.

The main topics currently under review are outlined below.

- Stratification
    - Seasonality / geography
    - Economics / geography
- Alternative estimation equations

- o Ownership information.
- Model structure
  - o Alternative heteroskedasticity assumptions
    - Gamma .5
    - Gamma .8
    - 2WLS
    - Seemingly Unrelated Regressors

This part of this study will make use of economic theory to formalize modeling strategies and the selection of optimal models. Statistical theory will be used to formalize the economic problem and solution as well as organize and present results.

Current imputation models use stratification to control for different seasonal patterns in regional electricity demand and weighted regression to control for heteroskedasticity. However, these technique assumes the same parameter estimates apply to each state in a given strata. Moreover, economic growth and seasonal patens are combined in these parameters. In addition to this, each strata-model combination is estimated independently of each other.

Consider as an alternative equations 1 and 2 below:

Equation 1

$$\begin{bmatrix} \bar{Y}_1 \\ \dots \\ \dots \\ \bar{Y}_n \end{bmatrix} = \begin{bmatrix} \bar{X}_1 & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \bar{X}_n \end{bmatrix} \times \begin{bmatrix} \bar{\beta}_1 \\ \dots \\ \dots \\ \bar{\beta}_n \end{bmatrix} + e$$

Equation 2

$$\begin{bmatrix} \bar{Y}_1 \\ \dots \\ \dots \\ \bar{Y}_n \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \dots \\ \dots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \bar{X}_1 & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \bar{X}_n \end{bmatrix} \times \begin{bmatrix} \bar{\beta}_1 \\ \dots \\ \dots \\ \bar{\beta}_n \end{bmatrix} + e$$

In equations one and two, n refers to each of the states in a given stratification group and each Y in $\bar{Y}_n$ refer to the reported data value(s) from each sampled firm that report to EIA-826. Equation 1 is no more than the matrix[7] representation of current imputation methods; equation 2 simply includes an intercept term. Both model representations are "fixed effects" models. With the second model, we could test if the intercepts are all equal to zero and if states' parameters are different among themselves. We are currently implementing the testing procedures to evaluate this assumption among others.

Consider equations 3 and 4 below.

Equation 3

$$e = \begin{bmatrix} e_1 \\ \dots \\ \dots \\ e_n \end{bmatrix} \approx N\left[\begin{pmatrix} 0 \\ \dots \\ \dots \\ 0 \end{pmatrix}, E\begin{pmatrix} e_1e_1' & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & e_ne_n' \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{nn} \end{pmatrix}\right]$$

[7] Running this regression in matrix form would produce the same outcomes as running each strata-equation separately.

Equation 4

$$e=\begin{bmatrix} e_1 \\ \dots \\ \dots \\ e_n \end{bmatrix} \approx N\left[\begin{pmatrix} 0 \\ \dots \\ \dots \\ 0 \end{pmatrix}, E\begin{pmatrix} e_1e_1' & \dots & \dots & e_1e_n' \\ e_2e_1' & \dots & \dots & e_2e_n' \\ \dots & \dots & \dots & \dots \\ e_ne_1' & e_ne_2' & \dots & e_ne_n' \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \dots & \dots & \sigma_{1n} \\ \sigma_{2n} & \dots & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}\right]$$

Equations 3 and 4 describe the distribution properties of the residuals from equations 2 and 4, and represent alternative assumptions in equations 1 and 2. The key difference between equations 3 and 4 is that in equation 3 the off-diagonal covariances e_je_k' (where j and k are different) are assumed to equal zero. That is, equation 3 does not allow for information feedback across models. Equation 4 does not make the same assumption.

To resolve this, the Seemingly Unrelated Coefficients (SUR) approach was used to include covariance information in equation 44 to estimate parameters efficiently. As a preliminary test, we instrumented SUR into currently used imputation methods with interesting results. These results have encouraged us to continue our research.

An additional step in our research is to replace existing strata with either FERC or Sub-FERC regions, summary results are forthcoming. Table one below outlines current stratification assumptions, FERC regions and sub-regions. We resolved for overlapping states with the use of population data as an instrumental variable. Note that we presented a different stratification scheme for residential consumption at the Spring 2005 meeting of the ASA Energy Committee.

Table 1

STATE	STRATA	NERC	SUBNERC
AK	AK	AK	ASCC
AL	SEA	SERC	SUTHERN
AR	SOU	SERC	ENTERGY
AZ	SWE	WSCC	AZNM
CA	WES	WSCC	CA
CO	SWE	WSCC	RMPA
CT	NEA	NE	ISONE
DC	NEA	MAAC	MAAC
DE	NEA	MAAC	MAAC
FL	SEA	FRCC	FRCC
GA	SEA	SERC	SUTHERN
HI	WES	HI	HI
IA	NEC	MAPP	MAPPUS
ID	NEW	WSCC	NWPP
IL	CEN	MAIN	MAIN
IN	CEN	ECAR	ECAR
KS	SOU	SPP	SPPN
KY	CEN	ECAR	ECAR
LA	SOU	SERC	ENTERGY
MA	NEA	NE	ISONE

MD	NEA	MAAC	MAAC
ME	NEA	NE	ISONE
MI	NEC	ECAR	ECAR
MN	NEC	MAPP	MAPPUS
MO	CEN	MAIN	MAIN
MS	SOU	SERC	SUTHERN
MT	NWC	WSCC	NWPP
NC	SEA	SERC	VARCAR
ND	NWC	MAPP	MAPPUS
NE	NWC	MAPP	MAPPUS
NH	NEA	NE	ISONE
NJ	NEA	MAAC	MAAC
NM	SWE	WSCC	AZNM
NV	WES	WSCC	NWPP
NY	NEA	NY	NY
OH	CEN	ECAR	ECAR
OK	SOU	SPP	SPPS
OR	NEW	WSCC	NWPP
PA	NEA	MAAC	MAAC
RI	NEA	NE	ISONE
SC	SEA	SERC	VARCAR
SD	NWC	MAPP	MAPPUS
TN	CEN	SERC	TVA
TX	SOU	ERCOT	ERCOT
UT	SWE	WSCC	NWPP
VA	SEA	SERC	VARCAR
VT	NEA	MAR	ISONE
WA	NEW	WSCC	NWPP
WI	NEC	MAIN	Main
WV	CEN	ECAR	ECAR
WY	NWC	WSCC	NWPP

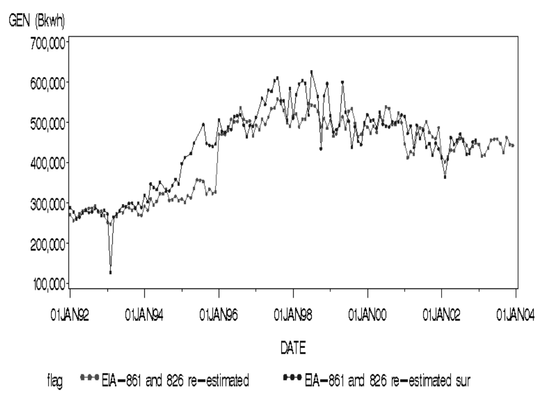
The preliminary findings of the first STRATA based SUR model can be subsetted in three groups.

- Estimation improvements;
- Structural changes; and
- Both estimation improvements and structural change.

Estimation improvements are those which performed better than both outlier detection models above. Structural changes are represented by those states for which the models broke down on or around the years of 2000, 2001 and 2002. Both, refers to those states that showed both cases at different points in time. We are currently cross-referencing the states were the models broke down to confirm or deny if these model's break-down could be indicative of structural change in the industry.

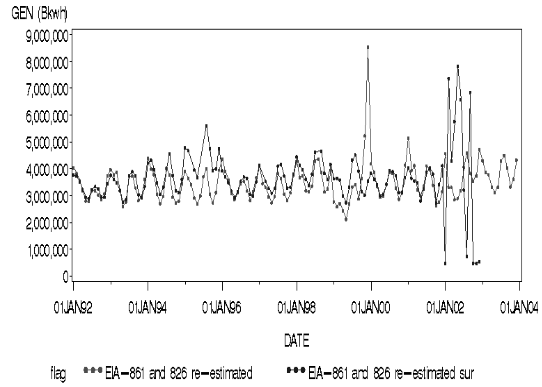
Graphs 7, 8 and 9 are examples of our findings. Graph 7 shows how a structural break in the original estimates was smoothed with SUR. Note that no outlier detection tools were used and that the break is present in published estimates and both alternative imputation methods. Moreover, the same sample and regressor data sets were used to estimate both lines as shown in the second graph. The use of the Drill-down tools shows how the brake is present in the original data.

Graphs 7: New Mexico Electricity Sold to the Industrial Sector



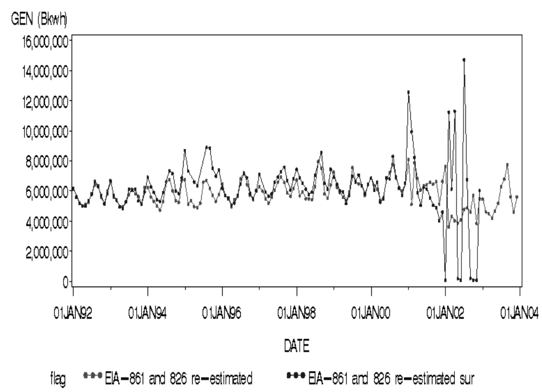
Graph 8 shows how outliers that could not be corrected in previous model runs were corrected for. The graph also shows how the model broke down in 2002. We are currently implementing outlier analysis tools into existing drill down systems to interpret the outliers.

Graph 8: New York Electricity Sold to the Residential Sector



Graph 9 also shows how the California model broke down in 2001.

Graph 9: California Electricity Sold to the Residential Sector



ACKNOWLEDGMENTS

James Knaub and Frederick Joutz provided useful comments and suggestions while Amy Jo Wheeler and Joel Douglas provided great analytical and editorial support. Thank you.

REFERENCES

- James Knaub extensive work can be found at http://interstat.statjournals.net/
- Applied Econometric Time Series, Walter Enders, Wiley and Mathematical Statistics
- Energy Information Administration, U.S. Department of Energy, Form EIA-861, http://www.eia.doe.gov/cneaf/electricity/forms/eia861/eia861instr.doc
- Energy Information Administration, U.S. Department of Energy, Form EIA-826, http://www.eia.doe.gov/cneaf/electricity/forms/eia826/eia826instr.doc
- Sampling techniques, Cochoran, Wiley and Mathematical Statistics-applied
- Survey Sampling, Theory and Methods, Arijit Chaudhuri & Horst Stenger